On the relevance of edge-conditioned convolution for GNN-based semantic image segmentation using spatial relationships

Patty Coupeau [1] Jean-Baptiste Fasquel [1] Mickaël Dinomais [1,2]

[1] University of Angers, LARIS - Systems Engineering Research laboratory

[2] Department of Physical Medicine and Rehabilitation - University Hospital of Angers

Séminaire des doctorants - 12 Mai 2022

Summary

1 Overview of my thesis

2 Context

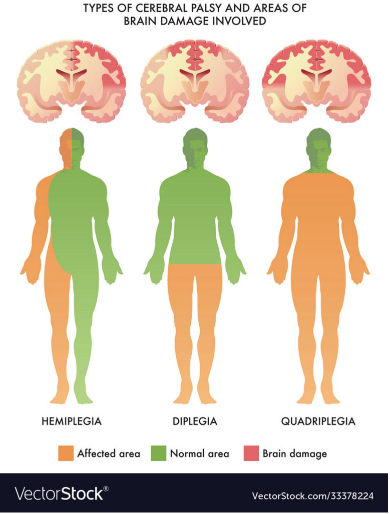
3 Method

4 Experiments

5 Conclusion and perspectives

Summary

1. Overview of my thesis
2. Context
3. Method
4. Experiments
5. Conclusion and perspectives



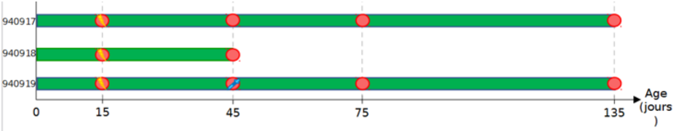
- Cerebral palsy is the leading cause of motor disability in children in France [a]
- Permanent motor disorder related to a non-progressive brain injury

- No existing treatment to "repair" the brain injury
- Pre-clinical trials using mesenchymal stem cells (MSCs)

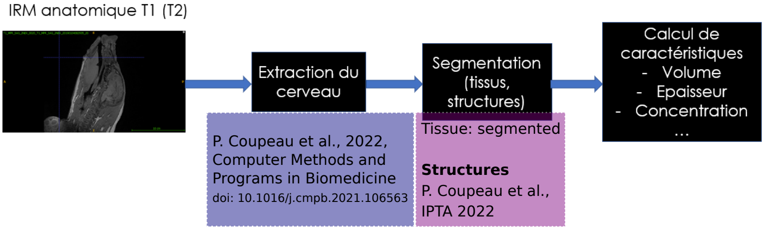
→ **Characterize the impact of early brain injury on brain development and the effects of MSCs**

[a] https://www.fondationparalysiecerebrale.org/

Pre-clinical study → animal model = piglet (similarity with the neonatal brain)
Macroscopic analysis → Magnetic Resonance Imaging



How to automate the MRI analysis pipeline in piglets? At different stages? Atlas-free



Basal ganglia (thalamus, putamen, caudate, pallidum): volume, structural organisation
→ place of interest in brain development → need for accurate segmentation

Summary

1 Overview of my thesis

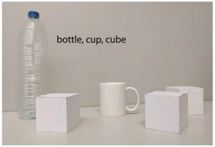
2 Context

3 Method

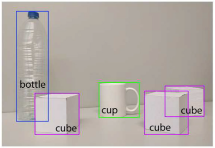
4 Experiments

5 Conclusion and perspectives

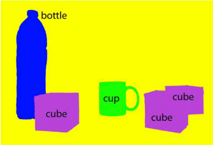
- Computer vision: many situations → convolutional neural network (CNN)
- Often ignored : relationships between entities → structural information
  - Spatial, photometric, geometric...



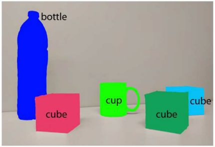
(a) Image classification



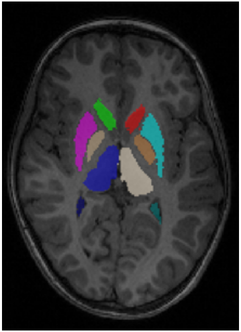
(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation



A. Garcia-Garcia, A survey on deep learning techniques for image and video semantic segmentation, Applied Soft Computing, 2018

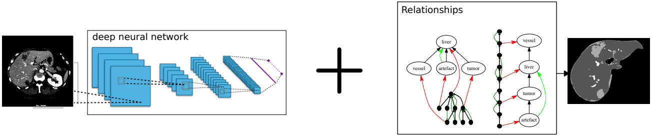
O. Duchenne et al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011

J. Zhou et al., Journal of Visual Communication and Image Representation, 2015

J.B. Fasquel et al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019

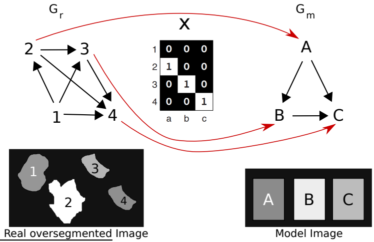
I. Bloch, Fuzzy sets for image processing and understanding, Fuzzy Sets and Systems, 2015

Preliminary semantic segmentation + structural information = refined segmentation



How to exploit structural information ?

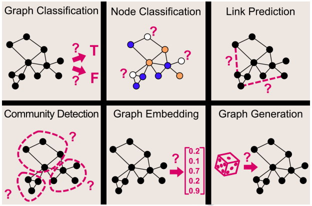
- Combinatorial optimization tools (constraint satisfaction, quadratic assignment problem)

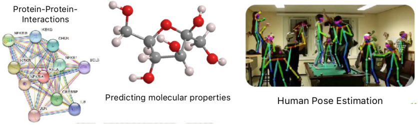


J. Chopin, J.B. Fasquel, H. Mouchere, R. Dahyot, and I. Bloch, 10th International Conference on Image Processing Theory, Tools and Applications,2020
J. Maciel and J.P.Costeira, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003
M. C. Vanegas, I. Bloch and J. Inglada, Fuzzy Sets and Systems, 2016

How to exploit structural information ?

- Graph neural network (GNN)





Constraints:

- Managing graphs of arbitrary size (depends on the CNN output)
- Managing both node and edge attributes

https://towardsdatascience.com/graph-convolutional-networks-deep-99d7fee5706f

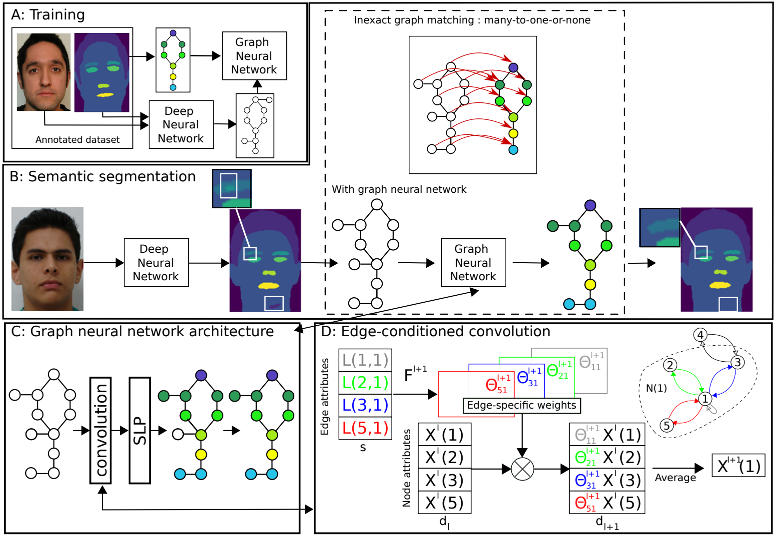
M. Fey, Workshop on Geometry and Machine Learning with Applications to Biomedical Engineering - University College London, 2020

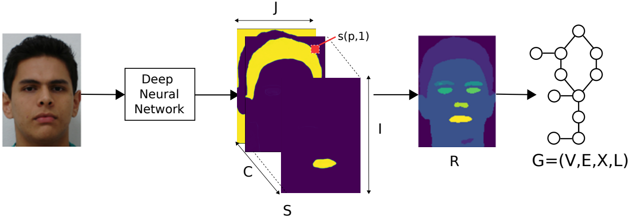
S. Ouyang et al., Remote Sensing, 2021

Q. Diao et al., Remote Sensing, 2022

Summary

1. Overview of my thesis
2. Context
3. Method
4. Experiments
5. Conclusion and perspectives


A: Training
Annotated dataset
Graph Neural Network
Deep Neural Network
B: Semantic segmentation
Deep Neural Network
Inexact graph matching : many-to-one-or-none
With graph neural network
Graph Neural Network
C: Graph neural network architecture
convolution
SLP
D: Edge-conditioned convolution
Edge attributes
L(1,1)
L(2,1)
L(3,1)
L(5,1)
s
F^{l+1}
Θ^{l+1}_{11} Θ^{l+1}_{21} Θ^{l+1}_{31} Θ^{l+1}_{51}
Edge-specific weights
Node attributes
$X^l(1)$
$X^l(2)$
$X^l(3)$
$X^l(5)$
d_l
$\Theta^{l+1}_{11}X^l(1)$
$\Theta^{l+1}_{21}X^l(2)$
$\Theta^{l+1}_{31}X^l(3)$
$\Theta^{l+1}_{51}X^l(5)$
d_{l+1}
Average
$X^{l+1}(1)$
N(1)



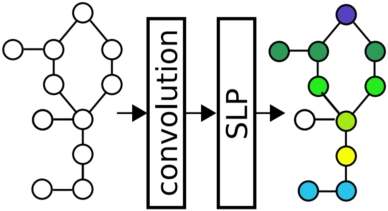
Segmentation map: $S \in R^{P \times C}$ from CNN

S(p, c) $\in [0, 1]$: probability of pixel p of belonging to class c

R: set of all resulting connected components

From R, construction of graph $G = (V, E, X, L)$

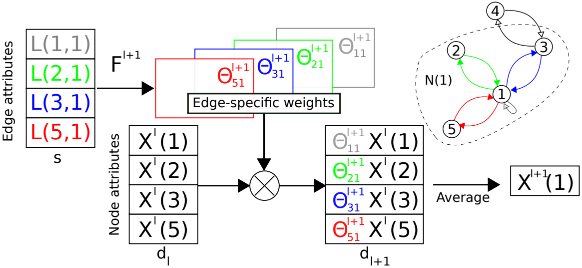
- V: set of nodes (each $v \in V$ corresponds to a region $R_v \in R$)
- E: set of edges
- $X : V \longrightarrow \mathbf{R}^C$: node attribute assignment function (average membership probability vector over the set of pixels $p \in R_v$)
- $L : E \longrightarrow \mathbf{R}^s$: edge attribute assignment function (depends on the considered spatial relationships)



Only 2 layers:

- convolution: aggregating neighborhood information related to each node (message passing)
- single layer perceptron (SLP): $\mathbf{R}^{d^{l+1}} \longrightarrow \mathbf{R}^{C}$, providing a class membership probability vector to each node of the graph

Edge-conditioned convolution: ECConv



$$X^{l+1}(i) = \frac{1}{|N(i)|} \sum_{j \in N(i)} F^{l+1}(L(j,i)) X^l(j) + b^{l+1} = \frac{1}{|N(i)|} \sum_{j \in N(i)} \Theta_{ji}^{l+1} X^l(j) + b^{l+1} \quad (1)$$

$F^{l+1} : \mathbf{R}^s \longrightarrow \mathbf{R}^{d^{l+1} \times d^l}$ mapping function (a multi-layer perceptron in our case)

X^{l+1} is computed using the average operator (permutation invariant operator)

M. Simonovsky et al., IEEE Conference on Computer Vision and Pattern Recognition, 2017

Summary

1 Overview of my thesis

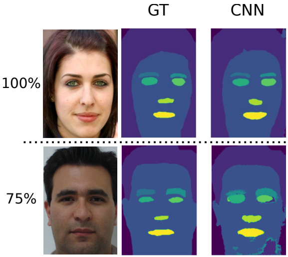
2 Context

3 Method

4 Experiments

5 Conclusion and perspectives

FASSEG-Instances

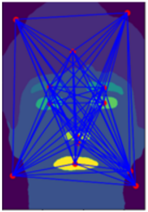


8 classes + background - 70 human faces

CNN: U-Net (splitting: 20/10/40)

Influence of dataset size (100% / 75%)

- Nodes: connected components ≥ 30 pixels
- Node attributes: membership probability vector of the region R_i
- Edge attributes: $L(i, j) = [d_{min}^{R_i, R_j}, d_{max}^{R_i, R_j}]$



https://github.com/Jeremy-Chopin/FASSEG-instances

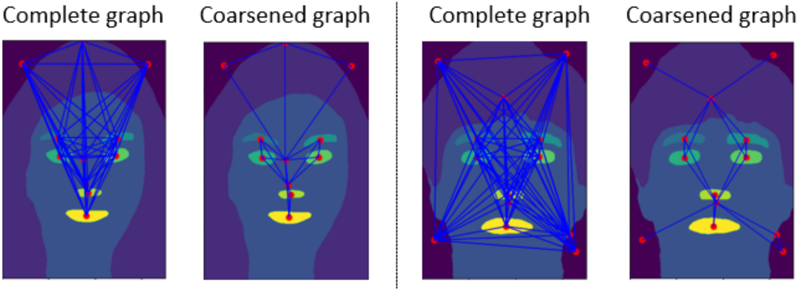
J. Chopin et al., 10th International Conference on Image Processing Theory, Tools and Applications, 2020

O. Ronneberger et al., Medical Image Computing and Computer-Assisted Intervention, 2015

Impact of the size of the neighborhood

Coarsened graph based on edge properties $L(i,j)$ $G_c = (V, E_c, X, L)$, where $E_c \subseteq E$

Hyperparameter radius ρ: limit distance between regions



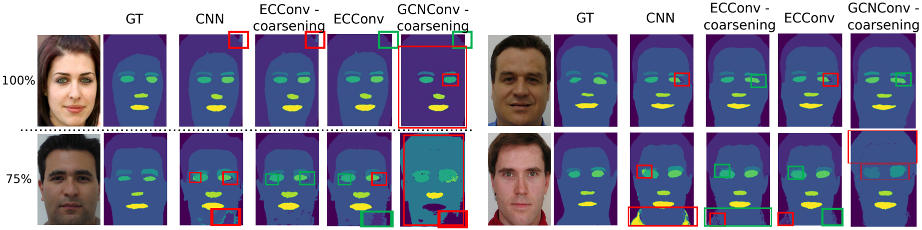
D. Bacciu et al., Neural Networks, 2020

Table: Graphs parameters for synthetic dataset and FASSEG. Values indicated are a mean over all images of the test dataset. Number of classes (C), of nodes ($|V|$) and of edges ($|E|$ and $|E_c|$), where $|E_c|$ is the number of edges after coarsening

Dataset	C	$\|V\|$	$\|E\|$	$\|E_c\|$
FASSEG 100%	9	12 (max: 26)	172 (max: 650)	33 (max: 134)
FASSEG 75%	9	17 (max: 86)	378 (max: 3867)	99 (max: 728)

Table: Segmentation results on FASSEG with CNN only and CNN followed by GNN (using ECConv or GCNConv). Complete graphs and coarsened ones are compared.

Method	75%			100%		
	DSC	***B-DSC***	***HD***	***DSC***	***B-DSC***	***HD***
CNN	0.798	0.675	54.40	0.845	0.745	27.20
ECConv	0.798	0.728	33.53	0.845	**0.769**	**19.76**
ECConv (G_c)	**0.804**	**0.731**	**32.00**	0.845	0.759	22.80
GCNConv* (G_c)	0.537	0.470	124.87	0.599	0.516	100.95



*GCNConv: does not consider edge attributes

T. Kipf et al., International Conference on Learning Representations, 2017

Summary

1 Overview of my thesis

2 Context

3 Method

4 Experiments

5 Conclusion and perspectives

- Coupling structural information and GNN : improves CNN-based semantic segmentation
- Relevance of ECConv i.e. consideration of both node (CNN output) and edge (spatial relationships) attributes
- Graph coarsening : improves robustness
- Simple architecture - Computation time : fast inference ($\leq$ 5s)

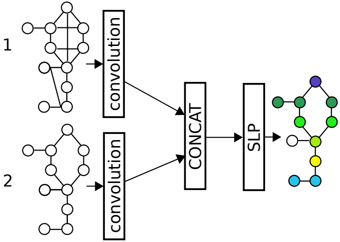
Preliminary experiments to be improved (larger datasets, GNN-architecture, etc.)

Perspectives

- Comparison with more recent CNN-based method: EfficientNet, CRF, U-Net Transformer
- Application to brain MRI for brain structures segmentation (3D, more complex)
- Improvement of the GNN architecture

Benefit of coarsening: multi-coarsening?

$\rightarrow$ Hyperparameter ρ ?

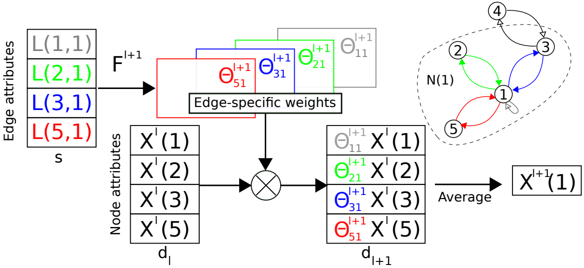


A. Hatamizadeh et al., IEEE Winter Conference on Applications of Computer Vision, 2022

B. Bhakti et al., IEEE Conference on Computer Vision and Pattern Recognition, 2020

S. Wang et al., IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, 2019

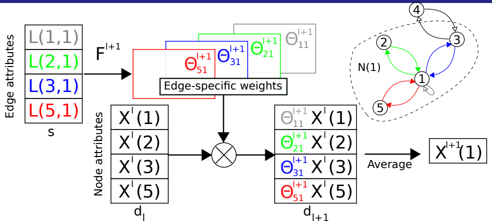
Thank you for your attention



For node $i \in V$, ECConv computes a new attribute $X^{l+1}(i)$ by combining different information from layer l :

- the attributes of the set $N(i)$ of nodes ($N(i) = \{j | (j, i) \in E\} \cup \{i\}$)
- the attributes of the set of related edges $\{L(j, i) | j \in N(i)\}$

M. Simonovsky et al., IEEE Conference on Computer Vision and Pattern Recognition, 2017



$$\begin{aligned} X^{l+1}(i) &= \frac{1}{|N(i)|} \sum_{j \in N(i)} F^{l+1}(L(j,i)) X^l(j) + b^{l+1} \\ &= \frac{1}{|N(i)|} \sum_{j \in N(i)} \Theta_{ji}^{l+1} X^l(j) + b^{l+1} \end{aligned} \tag{2}$$

$F^{l+1} : \mathbf{R}^s \longrightarrow \mathbf{R}^{d^{l+1} \times d^l}$ mapping function (a multi-layer perceptron in our case)

X^{l+1} is computed using the average operator (permutation invariant operator)

Dimensions of node attributes d^l ($l > 0$) are hyperparameters

Several convolution layers could be cascaded (only one in this study)

Nvidia Quadro RTX 3000 GPU - PyTorch libraries (torch_geometric.nn)

- optimizer: Adam
- loss function: negative log likelihood
- initial learning rate $lr_0 = 0.01$, reduction factor $\sigma = 5e - 4$

Synthetic

- 250 epochs
- $d1$=6
- train: 70 / test: 30

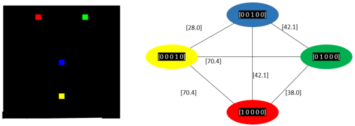
FASSEG-Instances

- 600 epochs
- $d1$=7
- train: 30 / test: 40

https://pytorch-geometric.readthedocs.io/en/latest/modules/nn.html

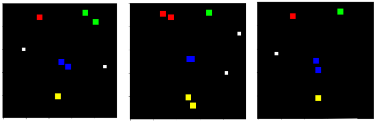
Synthetic

Reference



G

Altered images



4 classes + background

100 altered images

- Node attributes: membership probability vector of the region R_i
- Edge attributes: distance between barycenters of the connected regions R_i and R_j $(L(i,j) = |b_i - b_j|)$

Table: Results of classification of synthetic data with different configurations of graphs and convolution operators.

Method	Accuracy
ECConv (G_c)	**1.00**
ECConv	0.98
GCNConv (G_c)	0.59
ECConv (no node attributes)	0.20

Table: Segmentation results provided by the CNN only and our proposal. Results are provided for each class (not the background): Hr (hair), Fc (face), L-br (left eyebrow), R-br (right eyebrow), L-eye (left eye), R-eye (right eye), nose and mouth.

	75%						100%					
Method	**CNN**			**Proposal**			**CNN**			**Proposal**		
Class	***DSC***	***B-DSC***	***HD***	***DSC***	***B-DSC***	***HD***	***DSC***	***B-DSC***	***HD***	***DSC***	***B-DSC***	***HD***
Hr	0.924	0.773	126.26	**0.925**	**0.841**	**86.15**	0.941	0.825	85.18	0.941	**0.838**	**73.54**
Fc	0.948	0.917	48.29	**0.949**	**0.960**	**25.06**	0.957	0.955	24.38	0.956	**0.965**	**19.17**
L-br	0.681	0.547	65.33	**0.686**	**0.617**	**30.19**	0.751	0.679	11.41	0.751	0.678	11.41
R-br	0.667	0.537	65.77	0.652	**0.599**	**42.44**	0.744	0.584	42.50	0.745	**0.653**	**21.10**
L-eye	0.783	0.670	36.47	**0.804**	**0.707**	**23.06**	0.865	0.740	19.88	0.865	**0.782**	**10.11**
R-eye	0.783	0.643	36.97	0.783	**0.681**	**29.30**	0.837	0.718	14.29	0.837	**0.750**	**8.27**
Nose	0.742	0.559	41.41	**0.771**	**0.662**	**10.14**	0.797	0.684	8.47	0.797	**0.697**	**7.18**
Mouth	0.859	0.752	14.69	0.858	**0.779**	**9.42**	0.867	0.770	11.46	0.867	**0.791**	**7.31**